CHAPTER 4 METHODOLOGY

4.1 Introduction

In this chapter, we will first describe the models specification, the use of discrete choice model in analyzing mobile phone subscribers' churning decision. Then we incorporate both the characteristics of individuals and the attributes of calling patterns into the models. Finally, a brief overview of the data source is illustrated at the end of this chapter.

4.2 Model Specification

4.2.1 The Logit Model

Logistic regression (referred to as logit model) is the method that we use in this paper to analyze the dichotomous dependent variable – 'Churning' (y). This binary response variable takes the value of 0 or 1. Churning is 1 if the response is 'yes', and 0 otherwise. Logistic response function resembles an S-shaped curve. This method is chosen to overcome the violation of linear regression assumption[1]. In logistic regression, the method of estimation is the maximum likelihood method and the mean is used to get the log likelihood value (-2LL). The probability P_i first increases slowly with an increase in X, then the increase accelerates and finally stabilizes, but does not go beyond 1. The probability that subscriber i = 1,2,..., N will churn is P_i= Probability (Churn=1), $0 \le P_i \le 1$, and is given by

---

[1] In logistic regression,

1. There is nolinear relationship between the Y mean and X. The population mean of the Y's at each level of X are not a straight line.
2. The variance of errors for a given value of the independent variable X_i is $P_i(1- P_i)$, which is in contrary to the homoscedasticity assumption.
3. The errors are not normally distributed.

Pedhazur (1997), *"Multiple Regression in Behavioral Research: Explanation and Prediction"*, Wadsworth Thomson Learning, 3rd Ed.

$$P_i(Y=1 | X_1 = x_1, \dots X_p = x_p) = \frac{e^{(\alpha + \beta_1 \chi_1 + \beta_2 \chi_2 + \dots + \beta_p \chi_p)}}{1 + e^{(\alpha + \beta_1 \chi_1 + \beta_2 \chi_2 + \dots + \beta_p \chi_p)}}$$

The probability of an event happening over the probability of an event not happening is called the odd ratio for the event.

$$\frac{P_i}{1 - P_i} = e^{(\alpha + \beta_1 \chi_1 + \beta_2 \chi_2 + \dots + \beta_p \chi_p)}$$

The logarithm of the odd ratio is called the logit. The logit transformation produces a model that is linear in the parameters. The range for the value of logit is between $-\infty$ to $+\infty$, making it more appropriate for regression fitting.[2] The estimated form of the above equation is

$$g(x_1, \dots, x_p) = \ln \frac{P_i}{1 - P_i} = \alpha + \beta_1 \chi_1 + \beta_2 \chi_2 + \dots + \beta_p \chi_p$$

We assume, $\chi_1 \dots \chi_p$ are the independent variables

4.2.2 Regression Models and Variables

Churning is a dependent variable. We consider the effects of individual characteristics and calling patterns on the determinants of churning probabilities. Dummy variables are created for the qualitative explanatory choices.

[2] Chartterjee, Hadi & Price, "Regression Analysis By Example", John Wiley & Sons, p322.

(a) Model 1: Demographic Characteristics

In Model 1, churning is first regressed against the demographic characteristics, which include sex, marital status, education, and occupation.

$$g(x_1, \ldots, x_p) = \alpha + \beta_1 \text{Male} + \beta_2 \text{Single} + \beta_3 \text{College} + \beta_4 \text{Income} + \beta_5 \text{Clerical} + \beta_6 \text{Technician} + \beta_7 \text{Sales} + \beta_8 \text{Professional} + \beta_9 \text{Manager} + \mu \quad (1)$$

(b) Model 2: Consumer Calling Behavior Variables

In Model 2, we regress churning against consumer calling behavior or calling pattern variables. The factors being considered are the length of time in subscribing mobile services (Tenure), the initial subscribing fees (Initial fees), the total usage of mobile service per day measured in minutes (Usage), and the total monthly bill amount (Bill). In addition to that, the qualitative attributes include the fixed-line phone at home (Fixed-home), fixed-line phone in the office (Fixed-office), the valued added services (SMS, Voicemail, IDD, and Roaming), the purpose of using mobile services (Personal), and the person who settle the monthly payment (Self-pay).

$$g(x_1, \ldots, x_p) = \alpha + \beta_1 \text{Fixed-home} + \beta_2 \text{ Fixed-office} + \beta_3 \text{Tenure} + \beta_4 \text{Initial-fees} + \beta_5 \text{Usage} + \beta_6 \text{SMS} + \beta_7 \text{ Voicemail} + \beta_8 \text{IDD} + \beta_9 \text{Roaming} + \beta_{10} \text{Bill} + \beta_{11} \text{ Personal} + \beta_{12} \text{ Self-pay} + \mu \quad (2)$$

(c) Model 3: Demographic Characteristics and Consumer Calling Behavior Variables

The base model (Model 1) has been extended to include the calling behavior variables in the full model (Model 3).

$$
\begin{aligned}
g(x_1, \ldots, x_p) = {} & \alpha + \beta_1 \text{Male} + \beta_2 \text{Single} + \beta_3 \text{College} + \beta_4 \text{Income} + \beta_5 \text{Clerical} \\
& + \beta_6 \text{Technician} + \beta_7 \text{Sales} + \beta_8 \text{Professional} + \beta_9 \text{Manager} \\
& + \beta_{10} \text{Fixed-home} + \beta_{11} \text{Fixed-office} + \beta_{12} \text{Tenure} \\
& + \beta_{13} \text{Initial-fees} + \beta_{14} \text{Usage} + \beta_{15} \text{SMS} + \beta_{16} \text{Voicemail} \\
& + \beta_{17} \text{IDD} + \beta_{18} \text{Roaming} + \beta_{19} \text{Self-pay} + \beta_{20} \text{Bill} + \beta_{21} \text{Personal} \\
& + \mu
\end{aligned}
\tag{3}
$$

We assume that consumers are rational, thus, they make choices that maximize their perceived utility. If the consumers are not satisfied with the services provided, they will churn to other service providers when the derived benefits out weight the costs of churning.

We expect consumers' demographic characteristics to have little effect on churning as long as consumers are well informed, thus making good initial product choice. However, these variables may have an indirect effect on subscribers' consumption and churning decision as customers with similar demographic characteristics may share some common preferences or perceptions. A common viewpoint on a particular service provider's network and service quality may be due to the sharing of same network system in a common coverage area where the subscribers live. Moreover, some users may be more inclined to switch operator than others due to a change in occupation and income. In addition to that, we believe calling behavior variables are significant in affecting users' churning decision as these variables may reflect users' satisfaction level. These variables can also be good predictors of the users' budgets for mobile communication.

Churning has been an important determinant that affects a firm's profitability. Therefore, we would like to identify the factors that influence subscribers' churning decision. Based on the ability of models in predicting the probability of churning and the coefficients, we will gain an overview of the relationship between churning and the explanatory variables[3]. By evaluating each set of variables being included in the model, we want to find out the marginal contribution of each variable to the subscribers' churning decision.

4.3 Data Source

In this paper, we use 501 sample data collected from market surveys at the various shopping centres in KL, Selangor and the Klang Valley. The primary dataset in this study is collected through face-to-face interviews or telephone interviews, from November 2002 through February 2003. The respondents were subscribers of mobile phone services. Subscribers were asked on their demographic background, calling behavior and history of using mobile services with previous service providers. A survey questionnaire is attached at the end of this report (Appendix 3). We restrict our analysis to only one time churn and single user account, at any point in time.

This survey is based on the five operators, namely Celcom(M) Sdn Bhd (Celcom), Digi Telecommunications Sdn Bhd (Digi), Maxis Communication Sdn Bhd (Maxis), Telekom Malaysia Berhad (TM Touch) and Time dotCom (Time Cel). Because the market shares of Mobikom and Celcom (010) analogue services are very small, we excluded them from our analysis.

[3] "In binary regressand models, goodness of fit is of secondary importance. What matters is the expected sign of the regression coefficients and their statistical significance", Gujaratti Damodar N (2003).

Information on the service providers is mostly from a secondary source as we faced huge difficulties in accessing the primary data. We retrieved these resources mainly from magazines, press releases, articles and financial statements. This is due to the confidentiality of information as firms worry that rivals will access and make use of these figures in direct marketing competition.

Having discussed the model and the data source, we are going to explore further the descriptive summary of the sample data in the subsequent chapter.